Coalitional games for abstract argumentation[1]

Elise BONZON [a], Nicolas MAUDET [b] and Stefano MORETTI [c]
[a] *LIPADE, Universite Paris Descartes - elise.bonzon@parisdescartes.fr*
[b] *LIP6, Universite Paris 6 - nicolas.maudet@lip6.fr*
[c] *LAMSADE, Universite Paris Dauphine - Stefano.Moretti@dauphine.fr*

Abstract. In this work we address the issue of the uncertainty faced by a user participating in multiagent debate. We propose a way to compute the relative relevance of arguments for such a user, by merging the classical argumentation framework proposed in [5] into a game theoretic coalitional setting, where the *worth* of a collection of arguments (*opinions*) can be seen as the combination of the information concerning the defeat relation and the preferences over arguments of a "user". Via a property-driven approach, we show that the Shapley value [15] for coalitional games defined over an argumentation framework, can be applied to resume all the information about the worth of opinions into an attribution of *relevance* for the single arguments. We also prove that, for a large family of (coalitional) argumentation frameworks, the Shapley value can be easily computed.

Keywords. Argumentation, Coalitional games, Multiagent systems

1. Introduction

A multiagent debate can be seen as a collective gathering of arguments, possibly constrained by the rules of a protocol, whose outcome will eventually be evaluated (for instance thanks to the formal tools provided by argumentation theory [5,8]). A central problem faced by agents contributing in such a multiagent debate is that they have to put forward arguments taking into account their own goals, but also how the audience (the other agents taking part in the debate) may receive their arguments, and also possibly whether the rules of the debate will allow them to put forward these arguments. Consider, for instance, the attitude of politicians participating to public debates: their choice to embrace arguments or opinions, often depends by factors like, for instance, the popularity of the arguments, the share of voters supporting those opinions, a degree of personnel satisfaction, the consensus generated by those opinions in an assembly or in a forum, the contiguity with a political position, etc.

This results in complex decision-making problem, where most of the parameters are likely to be uncertain: what are the arguments known by other agents? what are their own goals? etc. Perhaps the most basic type of uncertainty is that it is virtually impossible to

[1] This work benefited from the support of the project AMANDE ANR-13-BS02-0004 of the French National Research Agency (ANR).

exactly predict what combination of arguments will result from the debate. In particular, even though full knowledge of the other agents' goals and arguments is assumed, the exact deadline of the debate may still be unknown, for instance. In this work, we thus investigate more specifically the uncertainty regarding the debate, and propose to exploit cooperative game theory to account for the decision-problem faced by an agent in this context.

We focus on the argumentation framework introduced in [5], where a set of arguments are represented by the nodes of a directed graph whose edges express the attacks or conflicts between pairs of arguments. Since the seminal paper [5], several authors have introduced and studied different notions of acceptability for arguments, assigning to each argument the status of 'accepted' or 'rejected', which is perhaps a simplistic manner to compare arguments in decision-making applications. Recently, few studies [1,3,4,11] attempted to evaluate intermediate levels of acceptability, using very different approaches. For example, in [11], a two-player strategic game between a 'proponent' and an 'opponent' is considered, where the strategies of the players are subsets of arguments, and the payoffs of the game are based on the structure of the directed graph associated to the argumentation framework.

Combining *preferences* and argumentation outcomes has been proposed initially in [14] and used in many work since. In this paper, we assume that the user has a cardinal preference relation over single arguments. From preferences over singleton arguments, it is possible to induce a preference order on the outcome of the debates (that is, sets of accepted arguments). For instance, I may prefer outcomes which make my favourite argument acceptable, regardless of whether many less preferred arguments are discarded. There are of course several ways to define such a preference ordering, and we discuss different ways to lift up preferences over subsets (*i.e.* coalitions) of arguments in our context. In fact, the way to combine the information about preferences and defeat relations within each coalition of arguments can be specified by the user according to different criteria, ranging from the pure and simple consideration of the exogenous user's preferences over opinions (disregarding the conflicts among arguments), to the other extreme where all arguments are indifferent for the user, and the notion of the worth of an opinion is exclusively based on the definition of argumentation semantics. Interestingly, in the latter case, the measure of importance for arguments provided by the Shapley value of the associated coalitional game, can be interpreted as a novel measure of acceptability, and can be compared with the existing notions from the related literature [1,3,4,11] (this comparison will be further discussed in future work). Intermediate coalitional argumentation frameworks, based on the combination of the users's preferences with a local (or myopic) observation of the abstract argumentation framework, are also considered.

The notion of *audience* is also not new: several authors have integrated this component in their favourite argumentation setting, see [2]. Basically, this means that the beliefs (or the preferences) of the other agents are considered. Recently, Grossi and van der Hoek [6] proposed an inspiring model, where agents are uncertain about the audience they face (the belief of the other agents). Based on this, they study how "adequate" are games, that is, roughly speaking, whether they are fair to both parties. Importantly, their approach takes into account the constraints imposed by the protocol. Our approach differs in that we address a related question (what is the best move to do) with a different tool —cooperative game theory— in order to compute the contribution of arguments to the satisfaction of the user.

To sum up, in this work, we thus introduce a framework for abstract argumentation, keeping into consideration the preferences over (collection of) arguments of an agent (hereafter, called "the user"), who deals with the problem of assessing the relevance of each argument with respect to her/his own objectives, and facing the uncertainty about which combination of arguments will result from the debate. The final goal of this paper, is then to measure the relative importance of arguments for the user, taking into account both her/his own preferences - as represented by a utility function defined over the set of arguments - and the information provided by the attack relations among arguments. In this direction, we merge the classical argumentation framework proposed in [5] into a game theoretic coalitional setting [13], where the "worth" of a *collection of arguments* (also called an *opinion*) can be seen as the combination of the information about the preferences of the user over arguments and the information concerning the conflicts between the arguments. In addition, classical power indices for coalitional games (in particular, the Shapley value [15]) are used to resume all the information about the worth of opinions into an attribution of *relevance* (or *priority*) for the single arguments.

The paper is organized as follows. We start in Section 2 with some basic definitions about abstract argumentation and coalitional games. Section 3 is devoted to the presentation of a *coalitional argumentation framework*, where a family of coalitional games is used to represent the worth of coalitions of arguments, and where the Shapley value of such games is re-interpreted as a measure of the relevance of arguments. Section 4 introduces a (myopic) coalitional argumentation framework, where the worth of an opinion is additive over non-attacked arguments, and the computation of the Shapley value is simple. In Section 5 we provide an axiomatic characterization of the Shapley value in the class of games introduced in Section 4.2. Section 6 deals with an extended notion of coalitional argumentation framework aimed to capture the uncertainty generated by an interaction protocol for multi-agent debates.

2. Basic notions and definitions

2.1. Argumentation framework

A *Dung Argumentation Framework* (DAF) is a directed graph $\langle \mathcal{A}, \mathcal{R} \rangle$, where the set of nodes $\mathcal{A}$ is a finite set of *arguments* and the set of arcs $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is a *binary defeat* (or *attack*) *relation* (i.e, $(i, j) \in \mathcal{R}$ means that argument $i \in \mathcal{A}$ attacks argument $j \in \mathcal{A}$). We say that a set of arguments $S \subseteq \mathcal{A}$ (also called an *opinion S*) attacks another opinion $T \subseteq \mathcal{A}$ in $\langle \mathcal{A}, \mathcal{R} \rangle$ if there exists $(s,t) \in S \times T$ with $(s,t) \in \mathcal{R}$, that is an attacks which originates from an argument in S and is directed against an argument in T. For each argument a we define the set of *predecessors* of a in $\langle \mathcal{A}, \mathcal{R} \rangle$ as the set $Pr(\mathcal{R}, a) = \{ j \in \mathcal{A} : (j,a) \in \mathcal{R} \}$, and the set of *successors* of a is denoted by $Su(\mathcal{R}, a) = \{ j \in \mathcal{A} : (a, j) \in \mathcal{R} \}$ (if clear from the context, we omit notation $\mathcal{R}$ in $Pr(i)$ and $Su(i)$).

The main goal of argumentation theory is to identify which arguments and opinion are rationally "acceptable" according to different notions of acceptability. Some of the most common ones are the following. An argument $a \in \mathcal{A}$ is said *acceptable w.r.t.* $S \subseteq \mathcal{A}$ iff $\forall b \in A$: if $(b,a) \in \mathcal{R}$, then $\exists c \in S$ such that $(c,b) \in \mathcal{R}$. A set $S \subseteq \mathcal{A}$ is said to be: (*conflict-free*) iff S does not attack itself; (*stable*) iff S is conflict-free and attacks every argument in $\mathcal{A} \setminus S$; (*admissible*) iff S is conflict-free and S attacks every argument in $T \setminus S$

that attacks S; (*preferred*) iff it is a maximal (w.r.t. $\subseteq$) admissible extension; (*complete*) iff $\forall a \in \mathcal{A}$, if a is acceptable w.r.t. S, then $a \in S$; (*grounded*) iff S is the minimal (w.r.t. $\subseteq$) complete extension.

2.2. Coalitional games

A *coalitional game*, also known as *characteristic-form game* or *Transferable Utility (TU) game*, is a pair (N, v), where N denotes a finite set of *players* and v is the *characteristic function*, assigning to each $S \subseteq N$, a real number $v(S) \in \mathbb{R}$, with $v(\emptyset) = 0$ by convention. If the set N of players is fixed, we identify a coalitional game (N, v) with the corresponding characteristic function v. A group of players $T \subseteq N$ is called a *coalition* and $v(T)$ is called the *worth* of this coalition. We will denote by $\mathcal{G}$ the class of all coalitional games. Let $\mathcal{C} \subseteq \mathcal{G}$ be a subclass of coalitional games. Given a set of players N, we denote by $\mathcal{C}^N \subseteq \mathcal{C}$ the class of coalitional games in $\mathcal{C}$ with N as set of players. We define the set Σ_N of possible linear orders on the set N as the set of all bijections $\sigma : N \to N$, where $\sigma(i) = j$ means that with respect to σ, player i is in the j-th position. For $\sigma \in \Sigma_N$, the *marginal vector* $m^\sigma(v) \in \mathbb{R}^N$ is defined by $m_i^\sigma(v) = v(\{j \in N : \sigma(j) \leq \sigma(i)\}) - v(\{j \in N : \sigma(j) < \sigma(i)\})$ for each $i \in N$.

A *one-point solution* (or simply a *solution*) for a class $\mathcal{C}^N$ of coalitional games is a function $\psi : \mathcal{C}^N \to \mathbb{R}^N$ that assigns a payoff vector $\psi(v) \in \mathbb{R}^N$ to every coalitional game in the class. The Shapley value $\phi(v)$ of a game (N, v) is then defined as the average of marginal vectors over all $|N|!$ possible orders in Σ_N ($|N|$ is the cardinality of the set N). In formula

$$\phi_i(v) = \sum_{\sigma \in \Sigma_N} \frac{m_i^\sigma(v)}{|N|!} \text{ for all } i \in N. \tag{1}$$

An alternative representation of the Shapley value for each $i \in N$ is by the formula

$$\phi_i(v) = \sum_{S \subseteq N \setminus i} \frac{1}{|N| \binom{|N|-1}{|S|}} (v(S \cup i) - v(S)). \tag{2}$$

We recall some nice properties of the Shapley value of a cost game v: *efficiency* (EFF), i.e. $\sum_{i \in N} \phi_i(v) = v(N)$; *symmetry* (SYM), i.e. if $v(L \cup \{i\}) = v(L \cup \{j\})$ for all $L \subset N$ such that $i, j \in N \setminus L$, then $\phi_i(v) = \phi_j(v)$; *dummy player property* (DPP), i.e. if $i \in N$ is such that $v(L \cup \{i\}) - v(L) = v(\{i\})$ for all $L \subseteq N$, then $\phi_i(v) = v(\{i\})$; *additivity* (ADD), i.e. $\phi(v) + \phi(w) = \phi(v + w)$ for each $v, w \in \mathcal{C}^N$. It is well known that the Shapley value is the only solution that satisfies these four properties on the class $\mathcal{C}^N$ [15].

3. Coalitional argumentation framework

We define a *Coalitional Argumentation Framework* (CAF) as a triple $\langle \mathcal{A}, \mathcal{R}, v \rangle$ where $\langle \mathcal{A}, \mathcal{R} \rangle$ is a DAF and v is a map assigning to each opinion $S \subseteq \mathcal{A}$ a number $v(S) \in \mathbb{R}$. The value $v(S)$ represents the *worth* of the opinion S for the user (for example, it could measure the success provided by opinion S according to a criterion specified by the user, e.g., the popularity of the opinion). We assume that for each $a \in \mathcal{A}$, the worth (or *utility*) of the singleton $\{a\}$ is given by a (cardinal) preference relation over $\mathcal{A}$.

In the following, we also assume that a CAF $\langle \mathcal{A}, \mathcal{R}, v \rangle$ satisfy the following condition of compatibility between the map v and the DAF $\langle \mathcal{A}, \mathcal{R} \rangle$: (c.1) if $a \in \mathcal{A}$ is such that $Pr(a) = Su(a) = \emptyset$ (a is not connected to other arguments in $\langle \mathcal{A}, \mathcal{R} \rangle$), then $v(S \cup \{a\}) = v(S) + v(\{a\})$ for each opinion $S \subseteq \mathcal{A}$; (c.2) if $a, b \in \mathcal{A}$ are such that $Pr(a) = Pr(b)$ and $Su(a) = Su(b)$ (i.e., a and b are symmetric in the DAF) and $v(\{a\}) = v(\{b\})$, then $v(S \cup \{a\}) = v(S \cup \{b\})$ for each opinion $S \subseteq \mathcal{A}$.

Given a CAF $\langle \mathcal{A}, \mathcal{R}, v \rangle$ (satisfying conditions (c.1) and (c.2) as well), we study the problem of providing a measure representing the relevance of arguments, taking into account both the structure of the DAF and the worth of opinions as measured by v. In this direction, we focus on properties that such a measure of relevance should satisfy.

For instance, the SYM[2] property introduced in Section 2, states that two symmetric players in the DAF should have the same relevance, provided that their worth as singletons is the same (differently stated, the relevance of an argument does not depend from its label). Analogously, rephrasing the notion of dummy player in a CAF, the DPP says that disconnected arguments in a DAF should receive as value of relevance precisely their worth as singletons. Still, the EFF property imposes un upper bound over the scale for measuring the relevance of arguments (precisely, the sum of the relevance values must be equal to $v(\mathcal{A})$). Finally, an interesting reinterpretation of the ADD property suggests that the sum of the relevance values measured over two distinct CAFs $\langle \mathcal{A}, \mathcal{R}, v_1 \rangle$ and $\langle \mathcal{A}, \mathcal{R}, v_2 \rangle$ sharing the same DAF (for instance, v_1 ande v_2 may represent the success of opinions over two distinct populations, like women vs. men), should be equal to the degree of admissibility of arguments measured on $\langle \mathcal{A}, \mathcal{R}, v_1 + v_2 \rangle$.

Example 1. *Consider two CAFs,* $\langle \{1,2,3\}, \{(1,2),(2,1)\}, v \rangle$ *and* $\langle \{1,2,3\}, \{(1,2),(2,1)\}, v' \rangle$ *(satisfying conditions (c.1) and (c.2) as well).*

Consider the CAF $\langle \{1,2,3\}, \{(1,2),(2,1)\}, \bar{v} = v + v' \rangle$*. By ADD and DPP, the Shapley value of argument* 3 *is* $\phi_3(\bar{v}) = v(\{3\}) + v'(\{3\})$*, and by SYM and EFF,* $\phi_1(\bar{v}) = \phi_2(\bar{v}) = \frac{1}{2}(v(\mathcal{A}) + v'(\mathcal{A}) - (v(\{3\}) + v'(\{3\})))$.

This specific interpretation of properties satisfied by the Shapley value, has the merit to better contextualize and support the use of the Shapley value as a meaningful measure of the relevance of arguments in a CAF. On the other hand, in order to compute such a measure, the nature of the worth of opinions, as represented by map v in a CAF, should be further specified. If we completely ignore the information provided by the directed graph in a CAF, and we assume that the worth of opinions is additive over the worth of the singletons (i.e., $v(S) = \sum_{i \in S} v(\{i\})$, for each $S \subseteq \mathcal{A}$), then it is easy to check that the Shapley value of each argument i equals $v(\{i\})$. On the other hand, the objective of this work is to apply the machinery of coalitional games to combine the user's preferences over arguments with the defeat relation of a DAF, and a possible approach to tackle this problem is illustrated in Section 4.

[2]Notice that, given a CAF $\langle \mathcal{A}, \mathcal{R}, v \rangle$, there may exist arguments a and b that are symmetric players in v (according to definition provided in Section 2) that do not necessarily satisfy the condition $Pr(a) = Pr(b)$, $Su(a) = Su(b)$ and $v(a) = v(b)$ introduced in condition $(c.2)$. In a similar way, there may exist arguments $a \in \mathcal{A}$ that are dummy players in v that do not satisfy condition $Pr(a) = Su(a) = \emptyset$.

4. Combining preferences and conflicts

As we previously mentioned, the definition of the worth of a set of arguments may depend on multiple factors. First of all, the combination of the information about preferences and the attack relation can vary from the unique consideration of the user's preferences over opinions (disregarding the conflicts among arguments), to the other extreme where all arguments are indifferent for the user, and the notion of the worth of an opinion is exclusively based on the definition of argumentation semantics.

We can thus see the level of consideration of the argumentation system in a scale, from no acknowledgement at all to a total consideration, as shown in Table 1 and discussed in the following sections.

Consideration of the argumentation framework:		
not at all (game v^o)	partial (game $\hat{v}$)	total (game v^*)
The worth of an opinion S is additive:		
over all arguments of S (Section 4.1)	over the non-attacked arguments in S (Section 4.2)	over the arguments in the grounded extension of the DAF restricted to S (Section 4.3)

Table 1. Levels of combinations of the preference relation and the DAF.

4.1. No consideration of the acceptability

We consider here a very basic example, where the user is only interested on his/her preference over arguments, and does not care about the attack relation existing between those arguments. Precisely, consider a CAF $\langle \mathcal{A}, \mathcal{R}, v^o \rangle$ where the worth of opinions is additive over the arguments, i.e.

$$v^o(S) = \sum_{i \in S} v^o(\{i\}),$$

for each $S \subseteq \mathcal{A}$ (by convention, $v^o(\emptyset) = 0$). So, the worth of an opinion S is measured as the sum of the worth of the single elements of S.

Example 2. *Consider the CAF $\langle \{1,2,3\}, \mathcal{R}, v^o \rangle$, such that the user's preference over the arguments is given by $v^o(\{1\}) = 0$, $v^o(\{2\}) = -1$, $v^o(\{3\}) = 1$. As the attack relation has no effect on the evaluation of the arguments, we do not need to precise it. The game v^o is provided in Table 2. The Shapley value of such a game is $\phi_1(v^o) = 0$, $\phi_2(v^o) = -1$ and $\phi_3(v^o) = 1$.*

S:	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
$v^o(S)$:	0	-1	1	-1	1	0	0

Table 2. The worth of each coalition $S \subseteq \{1,2,3\}$ in v^o.

As the Shapley value of each argument is the same as its cardinal utility, this game has no particular interest. However, taking into account the constraints induced by a protocol (as discussed in Section 6) could already make such a setting non-trivial.

4.2. Partial consideration of acceptability

We now focus on the case where the argumentation system is *partially* taken into account. As a particular example, we define here a notion of the worth of opinions combining the preferences of the user over single arguments and a "local" information about the attacks. Precisely, consider a CAF $\langle \mathcal{A}, \mathcal{R}, \hat{v} \rangle$ where the worth of opinions is additive over non-attacked arguments, i.e. if an opinion $S \subseteq \mathcal{A}$ forms, $\hat{v}(S)$ denotes the sum of the worth of single arguments that are not attacked within opinion S. Let $F_S = \{i \in S : \{i\}$ is not attacked by $S \setminus \{i\}\}$ be the set of non-attacked arguments in S, then

$$\hat{v}(S) = \sum_{i \in F_S} \hat{v}(\{i\}), \tag{3}$$

for each $S \subseteq \mathcal{A}$ (by convention, $\hat{v}(\emptyset) = 0$). So, the worth of an opinion S is measured as the sum of the worth of those single elements of S that are not attacked in S. We denote by $\hat{\mathcal{V}}^{\mathcal{A}}$ the class of CAFs $\langle \mathcal{A}, \mathcal{R}, \hat{v} \rangle$ on $\mathcal{A}$ introduced in this section, and by $\mathcal{G}\hat{\mathcal{V}}^{\mathcal{A}}$ the class of corresponding games defined by relation (3).

Example 3. *Consider the CAF* $\langle \{1,2,3\}, \{(1,2),(2,3)\}, \hat{v} \rangle$*, such that the preference over each argument i is the same and is equal to* 1*. The game* $\hat{v}$ *is provided in Table 3. The Shapley value of such a game is* $\phi_1(\hat{v}) = \phi_3(\hat{v}) = \frac{1}{2}$ *and* $\phi_2(\hat{v}) = 0$*. So the greatest relevance is given to arguments* 1 *and* 3 *(note that the opinion* $\{1,3\}$ *is the only stable one).*

S:	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
$\hat{v}(S)$:	1	1	1	1	2	1	1

Table 3. The worth of each coalition $S \subseteq \{1,2,3\}$ in $\hat{v}$.

Alternatively, suppose that the argument 2 *is preferred to the other ones, and thus the worth of opinion* 2 *is, for instance,* $\hat{v}(\{2\}) = 1$*, and* $\hat{v}(\{1\}) = \hat{v}(\{3\}) = 0$*. Now the relevance assigned by the Shapley value is* $\frac{1}{2}$ *to arguments* 2*,* 0 *to argument* 3*, and* $-\frac{1}{2}$ *to argument* 1*: the user would be worse off by attacking the most beneficial and non-defended argument* 2*, whereas she/he would receive no detriment in adopting argument* 3*.*

In general the Shapley value is hard to calculate, since it requires a number of operations that is exponential in the number of arguments. However, for the specific class of CAFs introduced in this section, it is possible to calculate the Shapley value easily. First, observe that a game $\hat{v}$ can be decomposed as the sum

$$\hat{v} = \sum_{i \in \mathcal{A}} \hat{v}(\{i\}) \hat{u}_{i,Pr(i)} \tag{4}$$

where, for each $i \in \mathcal{A}$ and each $S \subseteq \mathcal{A} \setminus \{i\}$,

$$\hat{u}_{i,S}(T) = \begin{cases} 1 \text{ if } i \in T \text{ and } S \cap T = \emptyset \\ 0 \text{ otherwise,} \end{cases} \tag{5}$$

for each $T \subseteq \mathcal{A}$.

Remark 1. *Consider a game $(\mathcal{A}, \hat{u}_{i,S})$ defined according to relation (5), with $i \in \mathcal{A}$ and $S \subseteq \mathcal{A} \setminus \{i\}$. Notice that $\langle \mathcal{A}, \mathcal{R}, \hat{u}_{i,S} \rangle$ is an element of $\hat{\mathcal{V}}^N$, with $\mathcal{R} = \{(j,i) | j \in S\}$, that is the set of attacks $\mathcal{R}$ is formed by the attacks of arguments in S to i.*

In order to prove that the Shapley value of a game $\hat{v}$ is easy to calculate, we need the following lemma.

Lemma 1. *The Shapley value of game $(\mathcal{A}, \hat{u}_{i,S})$, with $i \in \mathcal{A}$ and $S \subseteq \mathcal{A} \setminus \{i\}$, is such that*

$$\phi_k(\hat{u}_{i,S}) = \begin{cases} \frac{1}{s+1} & \text{if } k = j, \\ \frac{1}{s(s+1)} & \text{if } k \in S, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

for each $i \in \mathcal{A}$ and where $s = |S|$ (with $s = 0$ if $S = \emptyset$).

Proof. By the DPP property, we immediately have that $\phi_k(\hat{u}_{i,S}) = 0$ for each $k \in \mathcal{A} \setminus (S \cup \{i\})$. By relation (1), the S¡hapley value of argument i is the ratio between the number of permutations $\sigma \in \Sigma_{\mathcal{A}}$ in which $\sigma(i) < \sigma(j)$ for each $j \in S$, denoted by $n(s+1)$, and the total number of permutations $a!$, with $a = |\mathcal{A}|$. However, it is easy to check that such a ratio does not depend on the number of arguments in $\mathcal{A} \setminus (S \cup \{i\})$ (to see this, it suffices to notice that for each fixed permutation in $\sigma^{S,i} \in \Sigma_{S \cup \{i\}}$ of elements in $S \cup \{i\}$, the total number of permutations in $\Sigma_{\mathcal{A}}$ that preserve the ordering $\sigma^{S,i}$ is a constant $K(n-s-1)$). Then, we have that $\phi_i(\hat{u}_{i,S}) = \frac{n(s+1)}{a!} = \frac{s!\ K(n-s-1)}{(s+1)!\ K(n-s-1)} = \frac{1}{s+1}$, where $s!$ is precisely the number of permutations $\sigma^{S,i} \in \Sigma_{S \cup \{i\}}$ in which $\sigma^{S,i}(i) = 1$. Finally, if $S \neq \emptyset$, by properties EFF and SYM it follows that $\phi_k(\hat{u}_{i,S}) = \hat{u}_{i,S}(\mathcal{A}) - \frac{1}{s(s+1)} = -\frac{1}{s(s+1)}$ for each $k \in S$. □

Proposition 1. *Consider a CAF $\langle \mathcal{A}, \mathcal{R}, \hat{v} \rangle \in \hat{\mathcal{V}}^{\mathcal{A}}$. Then the Shapley value of game $(\mathcal{A}, \hat{v})$ is*

$$\phi_i(\hat{v}) = \frac{v(\{i\})}{|Pr(i)|+1} - \sum_{j \in Su(i)} \frac{v(\{j\})}{|Pr(j)|(|Pr(j)|+1)}, \tag{7}$$

for each $i \in \mathcal{A}$.

Proof. The proof immediately follows by relation (4) and the ADD property. □

Note that the Shapley value of an argument i in game $\hat{v}$ does not depend only on the number of predecessors (attackers) an argument has, but also on the number of successors (arguments attacked by i), and on the number of other attackers of the arguments attacked by i. An axiomatic characterization of the Shapley value on the class of the CAFs presented in this section is discussed in Section 5.

Example 4. *Consider a CAF $\langle \mathcal{A} = \{1,2,3,4,5,6\}, \mathcal{R} = \{(1,2),(2,1),(2,3),\ (3,4),\ (5,3),(6,5)\}, \hat{v} \rangle$, and such that $\hat{v}(\{i\}) = 1$ for each $i \in \mathcal{A}$. Using relation (7), we have that the Shapley value of game $\hat{v}$ is $(0, -\frac{1}{6}, -\frac{1}{6}, \frac{1}{2}, \frac{1}{3}, \frac{1}{2})$. Hence, $\phi_4(\hat{v}) = \phi_6(\hat{v}) > \phi_5(\hat{v}) > \phi_1(\hat{v}) > \phi_2(\hat{v}) = \phi_3(\hat{v})$. As in Example 3, arguments with the highest Shapley value are conflict free.*

For a comparison, we recall here that the ranking provided by the method introduced in [3,4] in the DAF $\langle \mathcal{A}, \mathcal{R} \rangle$ is $6 \succ 4 \succ 1 \sim 2 \succ 5 \succ 3$, whereas the ranking provided by the method in [11] is $6 \succ 1 \sim 2 \sim 4 \succ 3 \succ 5$.

4.3. Total consideration of acceptability

It is also possible to take into account the information provided by a DAF in a more explicit way, for example by means of the definition of a coalitional game based on the acceptability semantics. For instance, we could consider the coalitional game

$$v^*(S) = \sum_{i \in E_S} v^*(\{i\}),$$

where E_S is the grounded extension in the argumentation framework $\langle S, \mathcal{R}_S \rangle$, where $\mathcal{R}_S = \{(i,j) \in \mathcal{R} : i,j \in S\}$ is the set of attacks between arguments in S.

Example 5. *Consider the CAF* $\langle \{1,2,3\}, \{(1,2),(2,3),(3,2)\}, v^* \rangle$*, such that the worth of each singleton opinion* $\{i\}$ *is the same and equal to* 1*. The game* v^* *is provided in the Table 4, together with the grounded extensions corresponding to each DAF* $\langle S, \mathcal{R}_S \rangle$*. The Shapley value of such a game is* $\phi_1(v^*) = \frac{7}{6}$*,* $\phi_2(v^*) = \frac{1}{6}$ *and* $\phi_3(v^*) = \frac{2}{3}$*. As expected, the higher Shapley value is for argument* 1*, which is not attacked in the DAF. Moreover, argument* 3*, which defends itself against argument* 2*, has a higher value than* 2*, which cannot defend itself against* 1*.*

S:	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
E_S:	$\{1\}$	$\{2\}$	$\{3\}$	$\{1\}$	$\{1,3\}$	$\emptyset$	$\{1,3\}$
$v^*(S)$:	1	1	1	1	2	0	2

Table 4. The grounded extension and the worth of each coalition $S \subseteq \{1,2,3\}$ in v^*.

Alternatively, suppose that the user likes argument 2*, dislikes argument* 3 *and is neutral with respect to argument* 1*. A possible choice of the worth of each singleton opinion representing this situation is* 0 *for argument* 1*,* 1 *for argument* 2 *and* -1 *for* 3*. In this new situation,* $\phi_1(v^*) = -\frac{1}{2}$*,* $\phi_2(v^*) = \frac{1}{2}$ *and* $\phi_3(v^*) = -1$*. Note that although argument* 1 *is neutral to the user, its relevance is negative since it can only do some harm to coalitions the user likes.*

5. An axiomatic characterization

In this section we provide an axiomatic characterization of a solution in the specific class of coalitional games arising from CAFs in $\hat{\mathcal{V}}^{\mathcal{A}}$, as introduced in Section 4.2. For this purpose, we define a *solution* as a map $\psi : \mathcal{G}\hat{\mathcal{V}}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}}$. An interesting property for an index is the following.

Axiom 1 (Equal Impact of Attack)**.** *Let* $\langle \mathcal{A}, \mathcal{R}, \hat{v} \rangle \in \hat{\mathcal{V}}^{\mathcal{A}}$ *and* $i,j \in \mathcal{A}$*, with* $i \neq j$*. Let* $\hat{v}_{ij}$ *be the game in the CAF* $\langle \mathcal{A}, \mathcal{R} \cup \{(i,j)\}, \hat{v}_{ij} \rangle \in \hat{\mathcal{V}}^{\mathcal{A}}$ *with* $\hat{v}_{ij}(\{k\}) := \hat{v}(\{k\})$ *for each* $k \in \mathcal{A}$*. A solution* $\psi : \mathcal{G}\hat{\mathcal{V}}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}}$ *satisfies the property of Equal Impact of Attack (EIA) iff*

$$\psi_i(\hat{v}) - \psi_i(\hat{v}_{ij}) = \psi_j(\hat{v}) - \psi_j(\hat{v}_{ij}).$$

Property of EIA states that when a new attack between two argument i and j is added to (or removed from) a CAF, then the relevance of the two arguments should be affected in the same way. Differently stated, this property says that a consequence of an attack should be detrimental for both arguments involved in the defeat relation, since an attacks always decreases the worth of coalitions containing the involved nodes. The following proposition shows that the Shapley value of a game corresponding to a CAF $\langle \mathcal{A}, \mathcal{R}, \hat{v} \rangle \in \hat{\mathcal{V}}^{\mathcal{A}}$ satisfies the EIA property.

Proposition 2. *The Shapley value satisfies the EIA property (in the class of games $\mathcal{G}\hat{\mathcal{V}}^{\mathcal{A}}$).*

Proof. $\langle \mathcal{A}, \mathcal{R}, \hat{v} \rangle \in \hat{\mathcal{V}}^{\mathcal{A}}$ and $i, j \in \mathcal{A}$, with $i \neq j$. Of course the interesting case is $(i, j) \notin \mathcal{R}$. Define the game $\hat{v}_{ij}$ as in Axiom 1. We have that

$$\begin{aligned}
&\phi_i(\hat{v}) - \phi_i(\hat{v}_{ij}) \\
&= \left(\tfrac{v(\{i\})}{|Pr(\mathcal{R},i)|+1} - \textstyle\sum_{k \in Su(\mathcal{R},i)} \tfrac{v(\{k\})}{|Pr(\mathcal{R},k)|(|Pr(\mathcal{R},k)|+1)} \right) \\
&- \left(\tfrac{v(\{i\})}{|Pr(\mathcal{R}\cup\{(i,j)\},k)|+1} - \textstyle\sum_{k \in Su(\mathcal{R},i)\cup\{j\}} \tfrac{v(\{k\})}{|Pr(\mathcal{R}\cup\{(i,j)\},k)|(|Pr(\mathcal{R}\cup\{(i,j)\},k)|+1)} \right) \\
&= \tfrac{v(\{j\})}{(|Pr(\mathcal{R},j)|+1)(|Pr(\mathcal{R},j)|+2)},
\end{aligned}$$

where the first equality follows by Proposition 1 and the fact that in the directed graph $\mathcal{R} \cup \{(i, j)\}$, argument i has one more successor, i.e. argument j, and argument j has one more predecessor, i.e. argument i (w.r.t. the directed graph $\mathcal{R}$). In a similar way, we have that

$$\begin{aligned}
&\phi_j(\hat{v}) - \phi_j(\hat{v}_{ij}) \\
&= \left(\tfrac{v(\{j\})}{|Pr(\mathcal{R},j)|+1} - \textstyle\sum_{k \in Su(\mathcal{R},i)} \tfrac{v(\{k\})}{|Pr(\mathcal{R},k)|(|Pr(\mathcal{R},k)|+1)} \right) \\
&- \left(\tfrac{v(\{j\})}{|Pr(\mathcal{R},j)|+2} - \textstyle\sum_{k \in Su(\mathcal{R},i)} \tfrac{v(\{k\})}{|Pr(\mathcal{R},k)|(|Pr(\mathcal{R},k)|+1)} \right) \\
&= \tfrac{v(\{j\})}{|Pr(\mathcal{R},j)|+1} - \tfrac{v(\{j\})}{|Pr(\mathcal{R},j)|+2} = \tfrac{v(\{j\})}{(|Pr(\mathcal{R},j)|+1)(|Pr(\mathcal{R},j)|+2)},
\end{aligned}$$

where the first equality follows by Proposition 1 and the fact that in $\mathcal{R}$ and $\mathcal{R} \cup \{(i, j)\}$ argument j has the same set of successors, but one more predecessor (w.r.t. the directed graph $\mathcal{R}$). □

We can now introduce the main result of this section.

Theorem 1. *The Shapley value is the unique solution that satisfies EFF, SYM, DPP, ADD and EIA properties on the class $\mathcal{G}\hat{\mathcal{V}}^{\mathcal{A}}$.*

Proof. The Shapley value satisfies EFF, SYM, DPP, ADD properties on each coalitional game, and by Proposition 2, it also satisfies EIA on the class $\mathcal{G}\hat{\mathcal{V}}^{\mathcal{A}}$.

Now, take a solution $\psi : \mathcal{G}\hat{\mathcal{V}}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}}$ that satisfies EFF, SYM, DPP, ADD and EIA properties. Consider a game $(\mathcal{A}, \hat{u}_{i,S})$ defined according to relation (5), with $i \in \mathcal{A}$ and $S \subseteq \mathcal{A} \setminus \{i\}$. We first prove that $\psi(\hat{u}_{i,S}) = \phi(\hat{u}_{i,S})$. The proof is by induction to $|S|$.

If $|S| = 1$, then S has a unique element, say $k \in \mathcal{A}$, $k \neq i$. For each $j \in \mathcal{A} \setminus \{i, k\}$, by the DPP we have that $\psi_j(\hat{u}_{i,\{k\}}) = 0$. By the EIA property $\psi_i(\hat{u}_{i,\emptyset}) - \psi_i(\hat{u}_{i,\{k\}}) = \psi_k(\hat{u}_{i,\emptyset}) - \psi_k(\hat{u}_{i,\{k\}})$. By DPP, we have that $\psi_i(\hat{u}_{i,\emptyset}) = 1$ and $\psi_k(\hat{u}_{i,\emptyset}) = 0$ for each $k \in \mathcal{A}$, $k \neq y$. By EFF and DPP, we have that $\psi_i(\hat{u}_{i,\{k\}}) + \psi_k(\hat{u}_{i,\{k\}}) = 0$. All these conditions together imply that $\psi_i(\hat{u}_{i,\{k\}}) = \frac{1}{2}$ and $\psi_k(\hat{u}_{i,\{k\}}) = -\frac{1}{2}$. So, $\psi(\hat{u}_{i,\{k\}}) = \phi(\hat{u}_{i,\{k\}})$.

Now let $m \in \mathbb{N}$, $m \geq 2$ and suppose that the $\psi(\hat{u}_{i,S}) = \phi(\hat{u}_{i,S})$ has been proved for every S with $|S| \leq m-1$. Consider a coalition $S' = S \cup \{k\}$, with $k \in \mathcal{A} \setminus (S \cup \{i\})$ and $|S'| = m$. By DPP, $\psi_j(\hat{u}_{i,S'}) = 0$ for each $j \in \mathcal{A} \setminus (S \cup \{i,k\})$. In addition, we have that

$$0 = \psi_i(\hat{u}_{i,S}) - \psi_i(\hat{u}_{i,S'}) - \psi_k(\hat{u}_{i,S}) + \psi_k(\hat{u}_{i,S'}) = \tfrac{1}{(s+1)} - \psi_i(\hat{u}_{i,S'}) + \psi_k(\hat{u}_{i,S'}), \quad (8)$$

where the first equality follows by the EIA property and the second equality follows by the induction hypothesis. Moreover,

$$0 = \textstyle\sum_{j \in S \cup \{i,k\}} \psi_j(\hat{u}_{i,S'}) = \psi_i(\hat{u}_{i,S'}) + (s+1)\psi_k(\hat{u}_{i,S'}), \quad (9)$$

where the first equality follows by EFF and DPP, and the second one by SYM. Combining relations (8) and (9), we have that $\psi_k(\hat{u}_{i,S'}) = \frac{1}{(s+1)(s+2)} = \phi_k(\hat{u}_{i,S'})$ and $\psi_i(\hat{u}_{i,S'}) = \frac{1}{s+2} = \phi_i(\hat{u}_{i,S'})$, which implies, by the application of the induction hypothesis, that $\psi(\hat{u}_{i,S}) = \psi_k(\hat{u}_{i,S})$ for every $i \in \mathcal{A}$ and $S \subseteq \mathcal{A} \setminus \{i\}$.

Using relation (4) and the additivity of ψ, we have that $\psi(\hat{v}) = \psi(\hat{v})$ for each $\hat{v} \in \mathcal{G}\hat{\mathcal{V}}^{\mathcal{A}}$, which concludes the proof.

□

6. Protocols for debates under uncertainty

In Sections 3, a property-driven approach has been used to support the adoption of the Shapley value as a measure of the relevance of arguments. Consequently, the relevance of an argument $i \in \mathcal{A}$ has been measured by its expected marginal contribution

$$\psi_i(v) = \sum_{A \subseteq \mathcal{A} \setminus \{i\}} p_i(S)(v(S \cup i) - v(S)), \quad (10)$$

where the probability distribution p_i is such that $p_i(S) = \frac{1}{|\mathcal{A}|\binom{|\mathcal{A}|-1}{|S|}}$ for each $S \subseteq \mathcal{A} \setminus \{i\}$. On the other hand, in abstract argumentation, probabilities over arguments and opinions may capture the uncertainty related to a specific argumentation framework [10,7], or the uncertainty related to a set of rules of an interaction protocol for multi-agent debates [12]. Therefore, it makes sense to consider "probabilistic" CAFs, where the information concerning the probability that opinions form is given *a priori*, and such probabilities are taken into account in the computation of the relevance via relation (10).

For example, consider a multi-agent interaction protocol where arguments are introduced by agents one after the other, respecting the protocol's rule according to which an argument can be introduced at a certain stage of the debate only if it attacks or it is attacked by another argument previously introduced. According to such a protocols, we can assume that it is not plausible that all coalitions could be potentially considered, since certain opinions will never form. Looking at the formula (1), this means that orders inducing the formation of impossible opinions should not be taken into account, and therefore should be deleted by the computation of the average marginal contribution over the permutation set. Of course, such a removal of orders induces a new probability distribution $p_i(S)$ in relation (10), for each $i \in \mathcal{A}$ and $S \subseteq \mathcal{A} \setminus \{i\}$: the index obtained in this way is the so-called "generalized Shapley value", introduced in [9].

Example 6. *Consider again the CAF of Example 3, and the multi-agent interaction protocol described in this section. According to such a protocol, opinion* $\{1,3\}$ *will never form. Then, in formula (1), orders inducing the formation of opinion* $\{1,3\}$ *should be removed. Precisely, those orders to be deleted are* $1,3,2$ *and* $3,1,2$*. It is easy to check that the generalized Shapley value of game* $\hat{v}$ *(i.e., the average marginal contribution of arguments over the remaining orders) is* $\frac{1}{4}$ *for argument* 1*,* $\frac{1}{2}$ *for argument* 2 *and* $\frac{1}{4}$ *for argument* 3*: notice that player* 2 *has now the highest relevance. The probabilistic value induced by the uncertainty generated from the interaction protocol has inverted the relative ranking of relevance over arguments, with respect to the CAF introduced in Example 3.*

However, the computation of the generalized Shapley value implies the consideration of all the orderings of the elements in $\mathcal{A}$, and therefore is computationally very expensive, even on small instances. A possible alternative could be the application of simpler probability distributions in relation (10). In this direction, one could think of using a uniform probability distribution over collections of feasible opinions (i.e., a notion of "generalized Banzhaf value"), or of applying methods aimed to approximate the expected marginal contribution over feasible coalitions. These aspects and other computational issues are discussed in a longer version of this manuscript.

References

[1] Amgoud L., Ben-Naim J. (2013) Ranking-Based Semantics for Argumentation Frameworks. In *Scalable Uncertainty Management*, LNCS, Springer, pp. 134-147.

[2] Bench-Capon T. J. M., Doutre S., Dunne P. E. (2007) Audiences in argumentation frameworks. *Artificial Intelligence*, 171(1), 42–71.

[3] Besnard P., Hunter, A. (2001) A logic-based theory of deductive arguments. *Artificial Intelligence*, 128, 203-235.

[4] Cayrol C., Lagasquie-Schiex M.C. (2005) Graduality in Argumentation. *Journal of Artificial Intelligence Research*, 23, 245-297.

[5] Dung P.M. (1995) On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *AI*, 77(2): 321-357.

[6] Grossi D., van der Hoek (2013). Audience-based uncertainty in abstract argumentation games. In Proc. of IJCAI. (pp. 31-39).

[7] Haenni R., Kohlas J., Lehmann N. (2000). Probabilistic argumentation systems. In Handbook of defeasible reasoning and uncertainty management systems (pp. 221-288). Springer Netherlands.

[8] Leite J., Martins J. (2011) Social Abstract Argumentation. In IJCAI, pp. 2287–2292.

[9] Loehman E, Whinston A (1976) A generalized cost allocation scheme. In: Lin SAY (ed) Theory and measurement of economic externalities. Academic, New York, pp 87–101

[10] Li H., Oren N., Norman T.J. (2012). Probabilistic argumentation frameworks. In Theories and Applications of Formal Argumentation (pp. 1-16). Springer Berlin Heidelberg.

[11] Matt P.A., Toni F. (2008) A game-theoretic measure of argument strength for abstract argumentation. In *Logics in Artificial Intelligence*, LNCS, Springer, pp. 285-297.

[12] McBurney P., Parsons S. (2001). Dialogue games in multi-agent systems. Informal Logic, 22(3).

[13] Owen G. (1995) Game theory, 3rd edn. Academic, San Diego. 1st edn 1968

[14] Rahwan Y., Larson K. (2008) Pareto Optimality in Abstract Argumentation. In Proc. of AAAI, pp. 150-155)

[15] Shapley L.S. (1953) A value for n-person games. In: Kuhn HW, Tucker AW (eds) Contributions to the theory of games II. *Annals of mathematics studies*, vol 28. Princeton University Press, Princeton, pp 307–317. Reprinted in: Roth AE ed (1988a), pp 31–40